

1/23

FIG. 1A

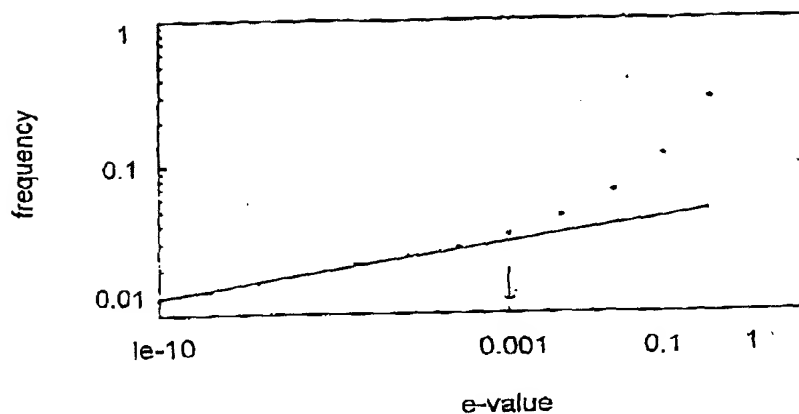


FIG. 1B

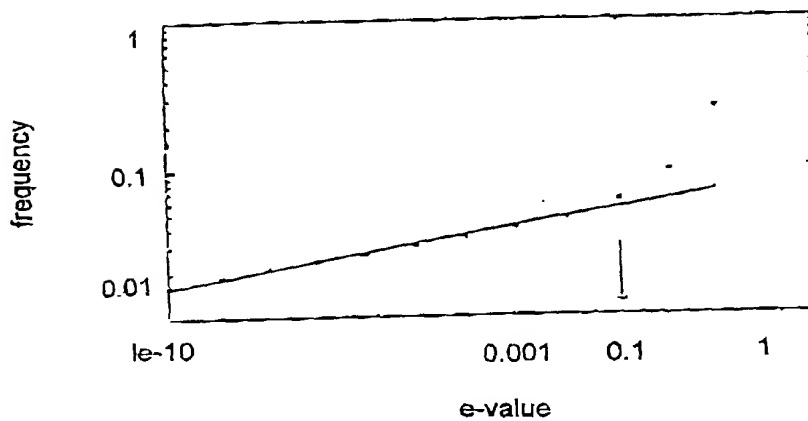
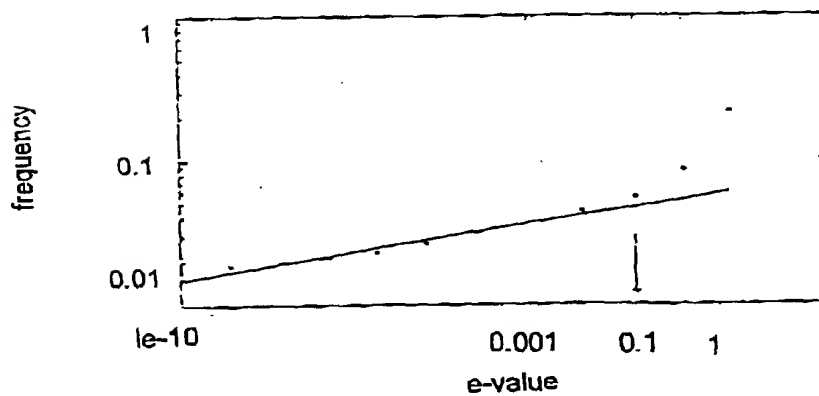
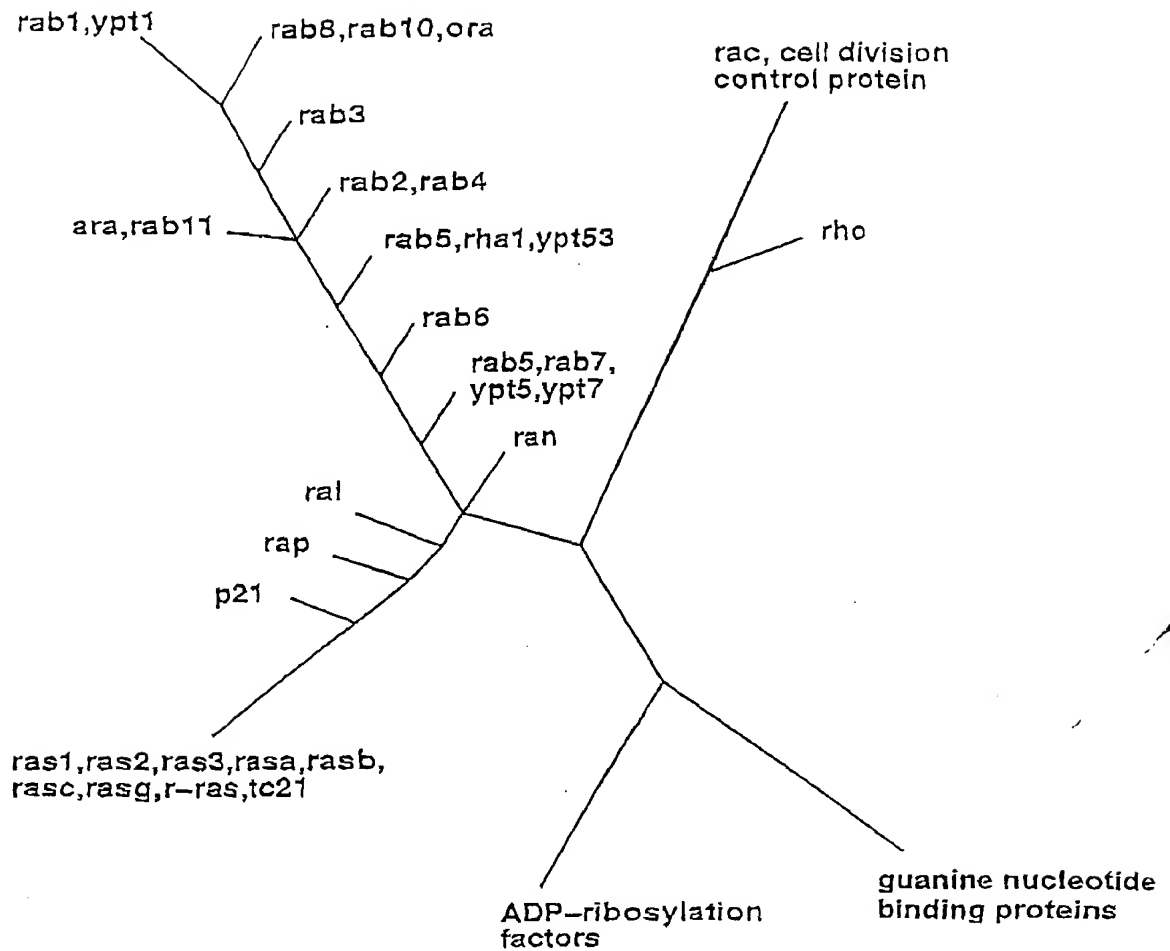


FIG. 1C



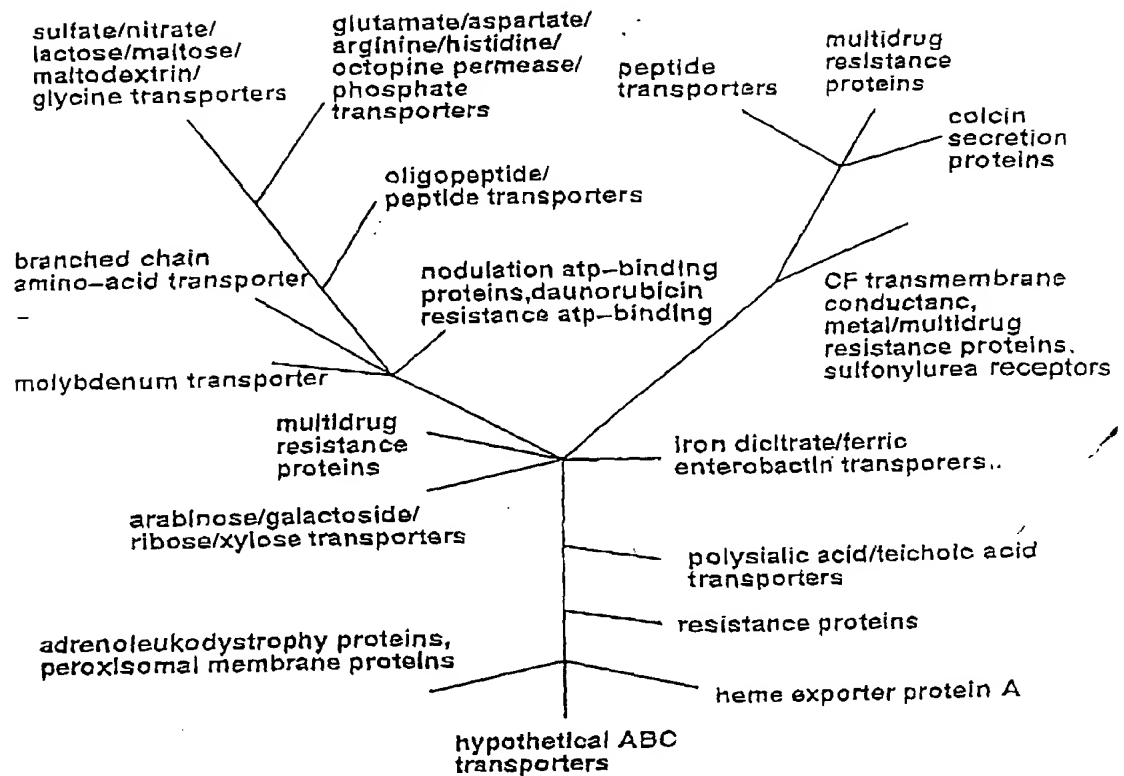
2/23

FIG. 2



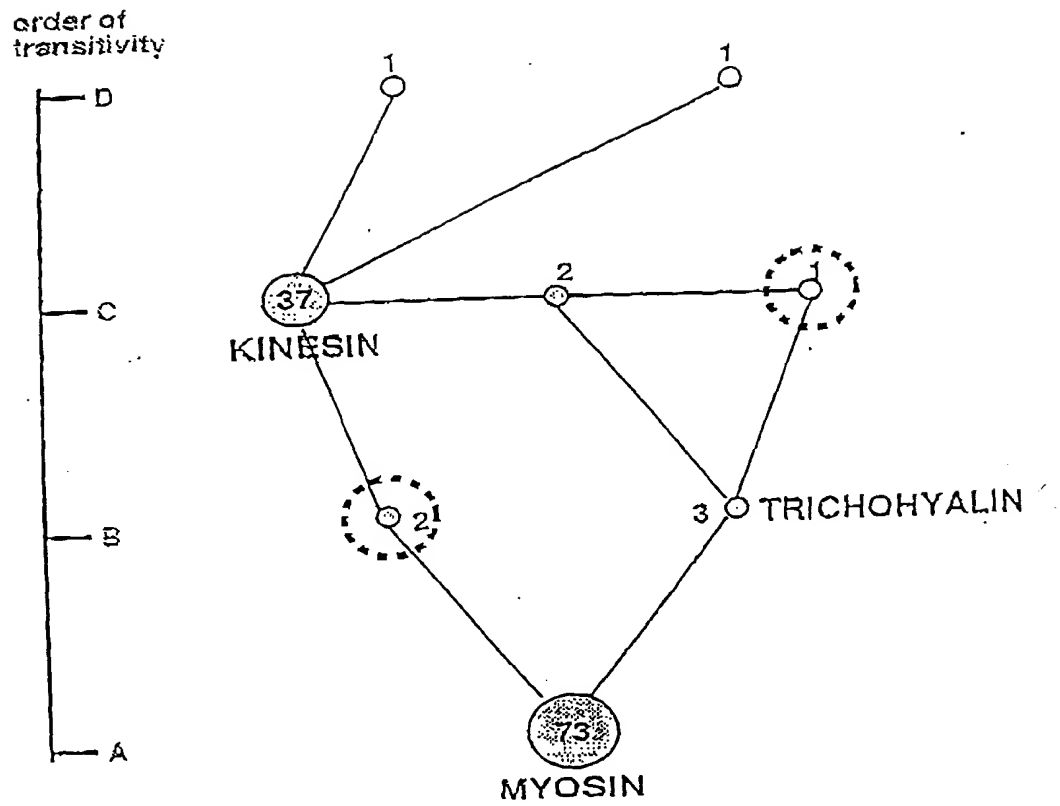
3/23

FIG. 3



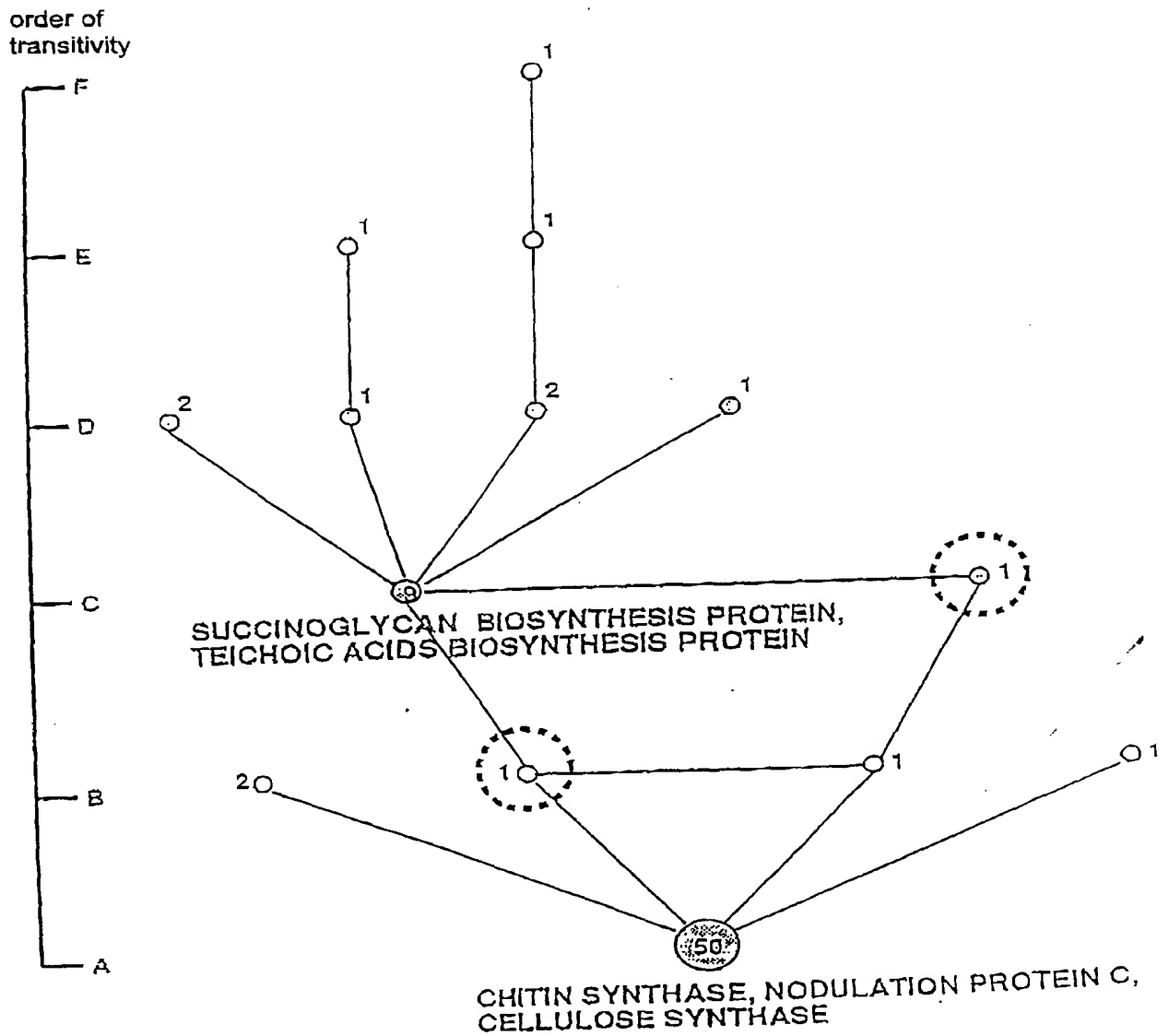
4/23

FIG. 4



5/23

FIG. 5



10220-8210960

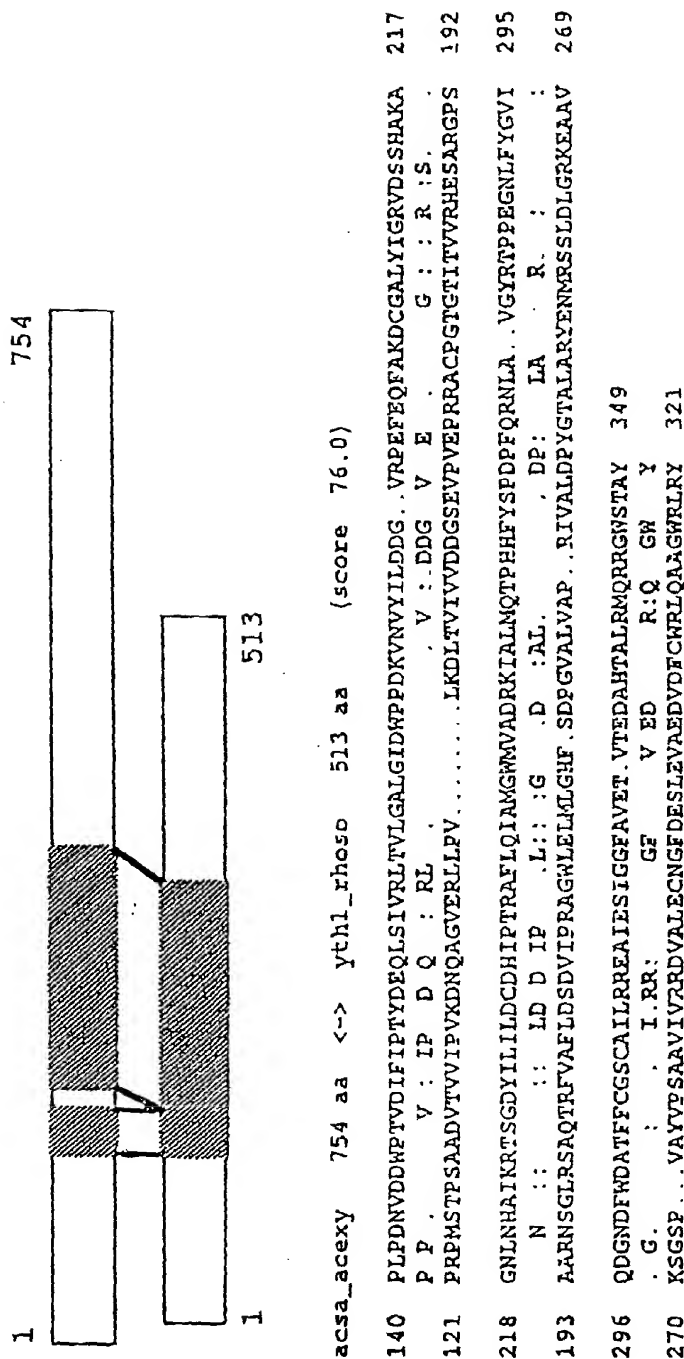
WO 99/39174

09/601278

PCT/IL99/00057

6/23

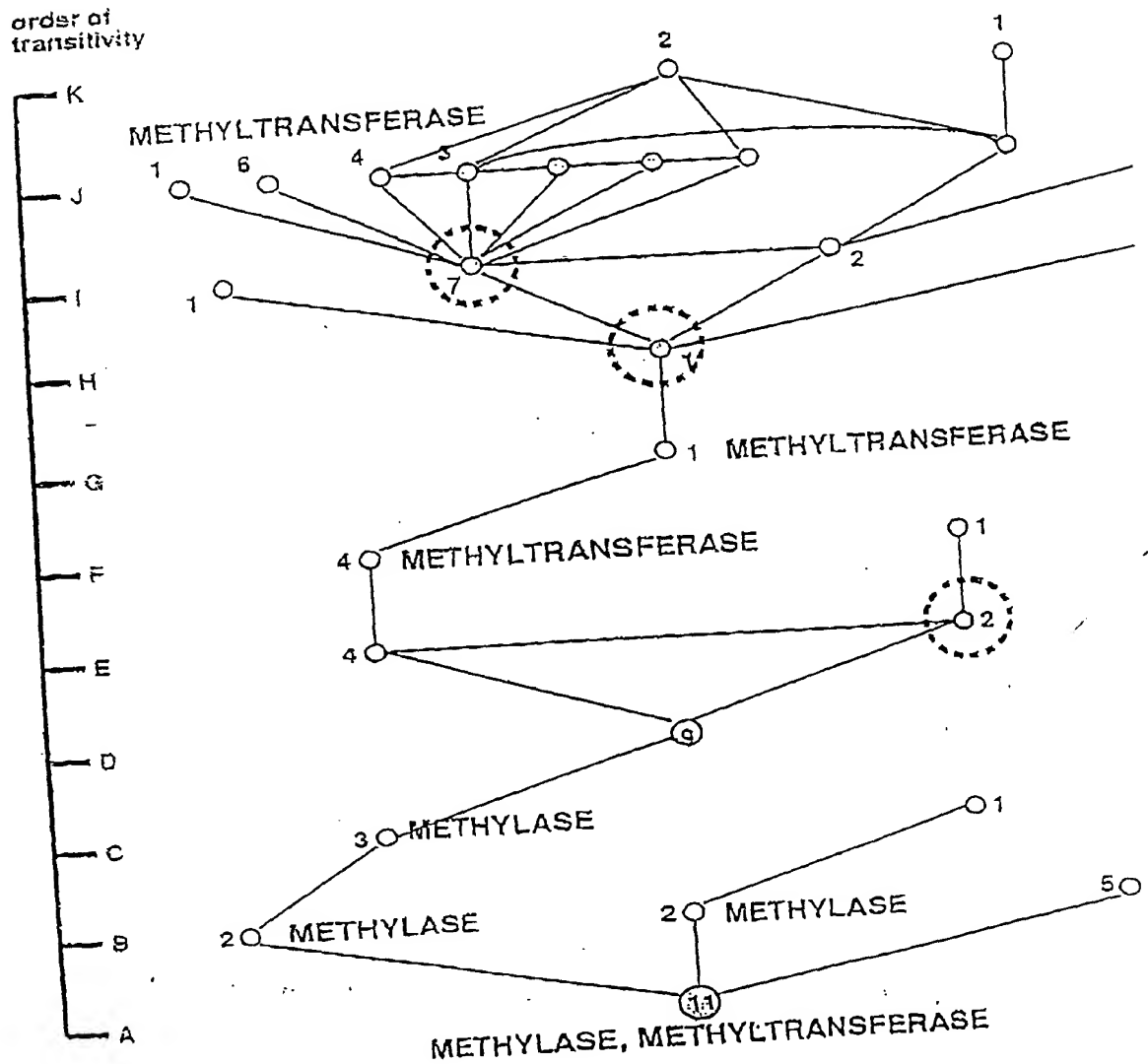
FIG. 6



alignment parameters: blosum62 matrix, gap penalties -12,-2

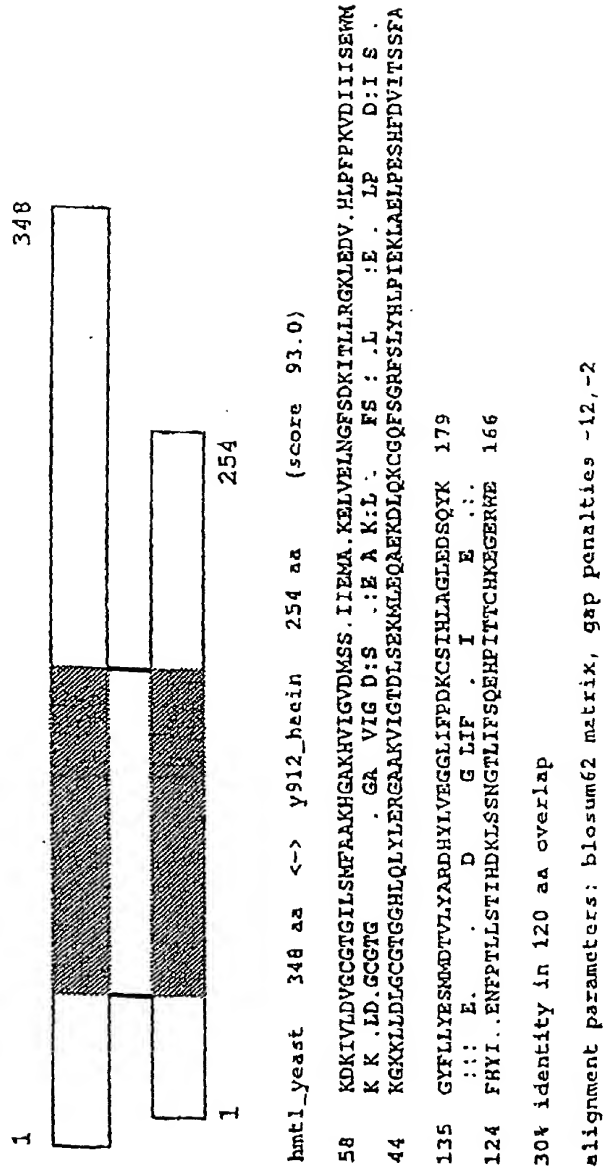
7/23

FIG. 7



8/23

FIG. 8





**FIG. 9**

Family	Size	True Positives	Family	Size Positives	True
HTH_LUXR_FAMILY.	36	0.42	C2_DOMAIN_2.	76	0.41
MICROBODIES_CTER.	79	0.39	GLU_CARBOXYLATION.	42	0.36
GRAM_POS_ANCHORING.	30	0.30	MDH.	42	0.29
PH_DOMAIN.	71	0.28	SH3.	136	0.28
EGF.	146	0.20	HMA.	24	0.17
ER_TARGET.	75	0.15	GATASE_TYPE_I.	75	0.09
C1Q.	22	0.09	SH2	125	0.09
B_KETOACYL_SYNTHASE.	37	0.05	SMALL_CYTOKINES_CC.	26	0.04
SIGMA54_INTERACT_1.	54	0.04	IGP5.	27	0.04
THYMIDYLATE_SYNTHASE.	28	0.04	OPSN.	51	0.02
DAG_PE_BINDING_DOMAIN.	62	0.00	TROPOMYOSIN.	40	0.00
KINESIN_MOTOR_DOMAIN.	38	0.00	PROTAMINE_P1.	38	0.00
PYRIDINE_REDOX_1.	37	0.00	ARF.	34	0.00
KRINGLE.	33	0.00	PROTEASOME_B.	31	0.00
PROTEASOME_A.	30	0.00	MYB_1.	27	0.00
TCP1_3.	27	0.00	PHYTOCHROME.	24	0.00
GLYCOSYL_HYDROL_F10.	23	0.00	PAIRED_BOX.	21	0.00
BAND_41_1.	20	0.00	CLOS_CELLULOSOME_RPT.	20	0.00

102220 82210950

WO 99/39174

PCT/IL99/00057

09/601218

10/23

FIG. 10

1514 clusters	mapped to	a single	new cluster	378 clusters	mapped to	few new	clusters	
	"Stable"		"Unstable"		"Stable"		"Unstable"	
The	Only new	Merged	Others	Perfect	Split, only	Split and	others	Total
same	sequences	with		split	new sequences	merged with		Stable
	added	singletons			added	singletons		
399	947	102	66	29	162	39	148	1678 (89%)

11/23

FIG. 11A

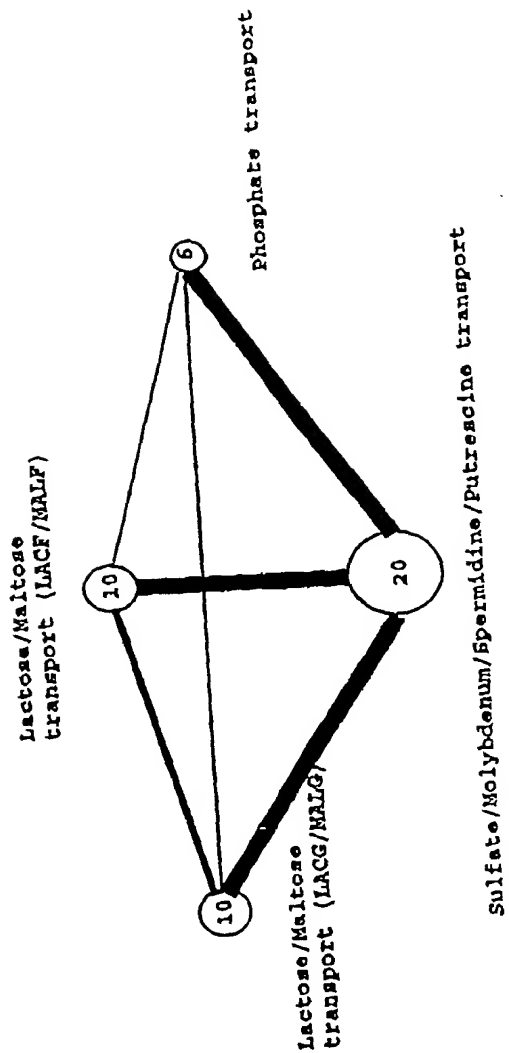
Cluster number	Size	Connection quality	Number of edges	Family
1643	5	0.29	219	B-cell antigen receptor complex associated protein
927	10	0.11	193	T-cell surface glycoprotein CD4
2613	3	0.03	20	polymeric-immunoglobulin receptor
5	326	0.01	226	immunoglobulins and major histocompatibility complex
1137	8	0.01	18	T-cell-specific surface glycoprotein CD28, cytotoxic T-lymphocyte protein
1189	8	0.01	9	myelin P0 protein precursor
1796	5	0.01	9	poliovirus receptor precursor

FIG. 11B

Cluster number	Size	Connection quality	Number of edges	Family
1831	5	0.38	248	T-cell receptor gamma chain C region
4	330	0.01	226	immunoglobulin V region
104	66	0.01	64	cell adhesion molecules, myelin-associated glycoprotein precursor
578	16	0.01	28	axonin-1 precursor, B-cell receptor CD22-beta precursor, and more
596	16	0.01	33	high/low affinity immunoglobulin epsilon/gamma FC receptor
856	11	0.01	11	recombination activating proteins, zinc finger, c3hcd type
1262	7	0.01	21	comifin (small proline-rich protein)
1636	5	0.01	8	T lymphocyte activation antigen CD80/CD86 precursor
1796	5	0.01	7	basigin precursor
				poliovirus receptor precursor

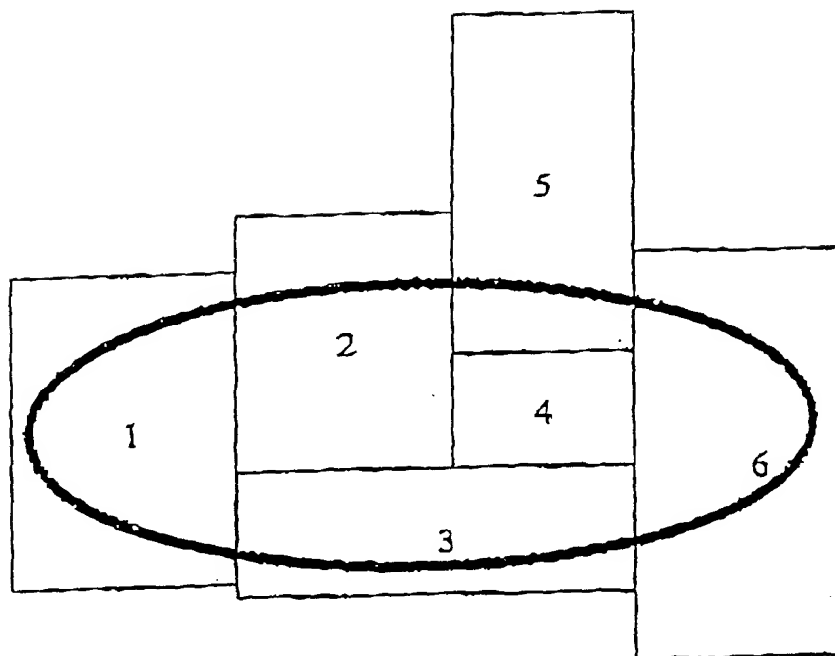
12/23

FIG. 12



13/23

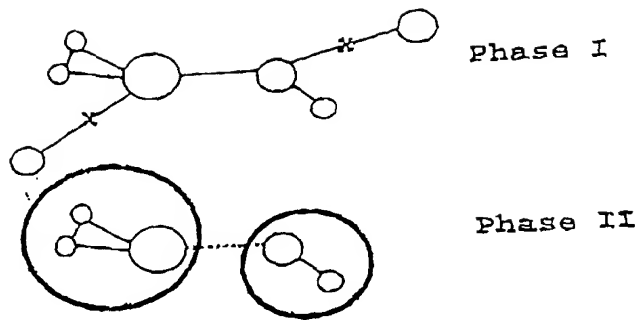
FIG. 13



SUBSTITUTE SHEET (RULE 26)

14/23

FIG. 14.



15/23

FIG. 15

Cluster number	Size	Order of transitivity	Family
1	718	2	protein kinases
2	593	2	globins
3	514	2	G-protein coupled receptors
4	330	2	immunoglobulin V region
5	326	2	immunoglobulins and major histocompatibility complex
6	318	2	homeobox
7	315	2	ribulose biphosphate carboxylase large chain
8	284	2	ABC transporters
9	260	1	zinc-finger C2H2 type
10	256	2	calcium-binding proteins
11	252	2	serine proteases, trypsin family
12	229	2	GTP-binding proteins - ras/ras-like family
13	221	2	myosin heavy chain, tropomyosin, kinesins
14	208	3	collagens, structural proteins
15	206	2	cytochrome p450

16/23

## CONTINUATION OF FIG. 15

Cluster number	Size	Order of transitivity	Family
16	198	2	GTP-binding elongation factors
17	196	2	tubulins
18	190	1	cytochrome b/b6
19	187	2	ATP synthases
20	172	2	heat shock proteins
21	171	2	alcohol dehydrogenases (short-chain)
22	171	2	snake toxins
23	152	2	NADH-ubiquinone oxidoreductase
24	142	2	bacterial regulatory components of signal transduction
25	141	3	DNA-binding proteins of HMG
26	140	1	nuclear hormones receptors
27	139	1	actins
28	139	1	intermediate filaments
29	138	2	GTP-binding, ADP-ribosylation factors family
30	136	1	neurotransmitter-gated ion-channels
31	133	2	zinc-containing alcohol dehydrogenases
32	133	2	cellular receptors, EGF-family
33	130	3	amylases
34	130	1	hemagglutinin
35	129	2	RNA-directed DNA polymerase
36	125	1	chaperones, chaperonins
37	122	2	phospholipase A2
38	120	2	insulins
39	115	1	cytochrome c
40	115	3	ketoacyl synthase
41	114	2	growth hormones (somatotropin, prolactin and related hormones)
42	113	1	glyceraldehyde 3-phosphate dehydrogenase
43	113	3	nuclear proteins, hn-RNP and sn-RNP, RNA processing proteins
44	110	1	viral nucleoprotein
45	109	1	cytochrome c oxidase subunit II
46	108	3	kazal serine protease inhibitors, secreted SPARC proteins
47	102	3	2Fe-2S ferredoxins, flavohemoproteins
48	102	2	viral genome polyproteins
49	102	1	developmental regulators - WNT family
50	101	1	cation transport ATPases



10220-8210960

WO 99/39174

09/601278

PCT/IL99/00057

17/23

FIG. 16

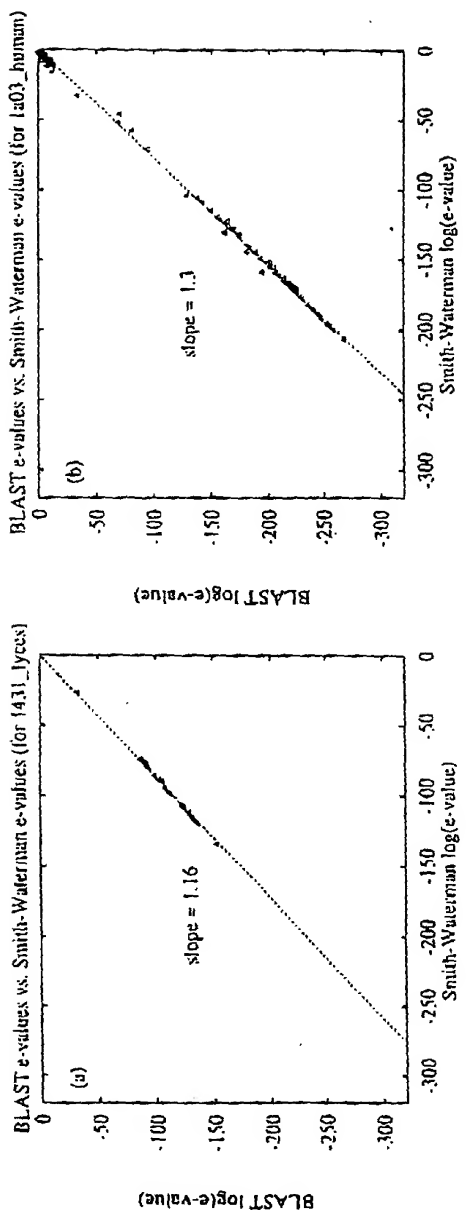
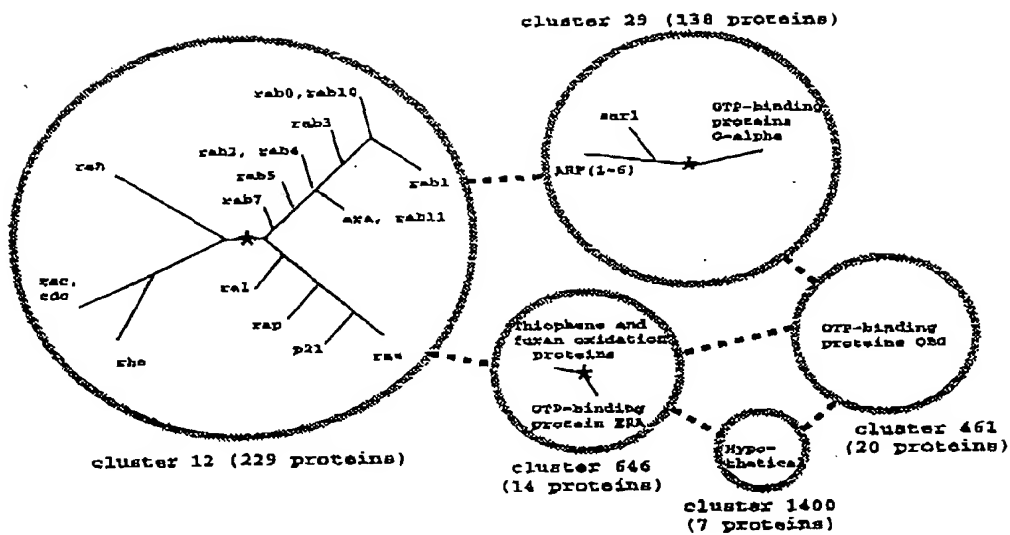


FIG. 17



1727

2294



20/23

FIG. 19

Confidence level	Cluster size							Total number of clusters
	over 100	51-100	21-50	11-20	6-10	2-5	1	
10 <sup>-100</sup>	8	18	90	234	528	3727	29870	34475
10 <sup>-95</sup>	8	19	100	240	537	3806	29086	33796
10 <sup>-90</sup>	8	20	111	256	545	3871	28224	33035
10 <sup>-85</sup>	8	23	119	262	563	4004	27189	32168
10 <sup>-80</sup>	8	25	133	264	594	4071	26140	31235
10 <sup>-75</sup>	9	31	132	275	623	4131	25051	30252
10 <sup>-70</sup>	10	34	138	293	653	4136	23943	29207
10 <sup>-65</sup>	11	32	156	309	660	4180	22911	28259
10 <sup>-60</sup>	13	34	171	319	677	4170	21772	27156
10 <sup>-55</sup>	15	40	178	334	676	4194	20646	26083
10 <sup>-50</sup>	15	51	184	350	676	4188	19463	24927
10 <sup>-45</sup>	17	53	197	362	696	4181	18282	23788
10 <sup>-40</sup>	21	54	203	383	714	4109	17129	22613
10 <sup>-35</sup>	23	53	213	393	780	4101	15801	21344
10 <sup>-30</sup>	26	53	232	415	774	4014	14428	19942
10 <sup>-25</sup>	29	57	252	421	788	3897	13191	18635
10 <sup>-20</sup>	32	64	263	436	779	3775	11839	17188
10 <sup>-15</sup>	35	64	270	464	808	3645	10620	15906
10 <sup>-10</sup>	38	76	293	457	802	3231	9112	14009
10 <sup>-5</sup>	51	92	315	431	884	2655	7169	11397
10 <sup>-0</sup>	51	94	315	456	703	2816	6167	10602

FOIA b 7 - DATED 8/22/09

WO 99/39174

09/601278

PCT/IL99/00057

21/23

FIGURE 20

Reference Database	True positives	Qsingle False positives	False negatives	True positives	Qset False positives	False negatives	Qset-relatives True positives	Upper-bound
PROSITE	68.4%	9.5%	22.1%	77.8%	11.1%	11.1%	75%	88.5%
Pfam	64.8%	7.3%	27.9%	76.7%	6.9%	16.4%	75%	83.1%

FIG. 21

Cluster no	Size	Family
23	152	
34	130	
44	110	
60	92	
67	86	
79	77	
85	74	
102	69	
110	64	
114	61	
120	58	
129	56	
132	55	
133	55	
135	54	
137	53	
138	53	
139	53	
142	52	
145	51	

09601278-022201

23/23

FIGURE 22

Cluster Number	Size	Family
4	330	Immunoglobulin V region
5	326	Immunoglobulins and major histocompatibility complex (MHC)
104	66	Cell adhesion molecules(CAM, n-CAM, ng-CAM, v-CAM, contactin, fascilin II), meylin associated glycoprotein, axonin-1 precursor, B-cell receptor CD22-beta precursor
578	16	High/low affinity immunoglobulin epsilon/gamma fc receptor
621	15	Interleukin-1 receptor, interleukin-1 binding protein, surface antigen
854	11	T-cell surface glycoprotein CD3 delta/epsilon/gamma chain precursor
927	10	T-cell surface glycoprotein CD4
1075	9	MHC class I NK cell receptor precursor
1137	8	T-cell-specific surface glycoprotein CD28, cytotoxic T-lymphocyte protein 4
1189	8	Meylin PO protein
1262	7	T lymphocyte activation antigen CD80/CD86 precursor
1301	7	Intercellular adhesion molecule precursor (ICAM)
1468	6	Hemagglutinin precursor
1636	5	Basigin precursor
1637	5	Probable cell adhesion molecule involved in regulating T-cell activation
1643	5	B-cell antigen receptor complex associated protein
1727	5	Interleukin-12 beta chain precursor
1796	5	Poliovirus receptor, $\alpha$ x-2 membrane glycoprotein
1831	5	T-cell receptor gamma chain C region
1938	4	T-cell surface antigen CD2 precursor
2294	3	Alpha-1b-glycoprotein
2613	3	Olymeric-immunoglobulin receptor
4622	1	Beta-2-microglobulin
4763	1	T-cell surface glycoprotein CD4
5186	1	Fasciclin III precursor
5583	1	Lymphocyte activation gene
5847	1	Immunoglobulin mu chain C region membrane bound